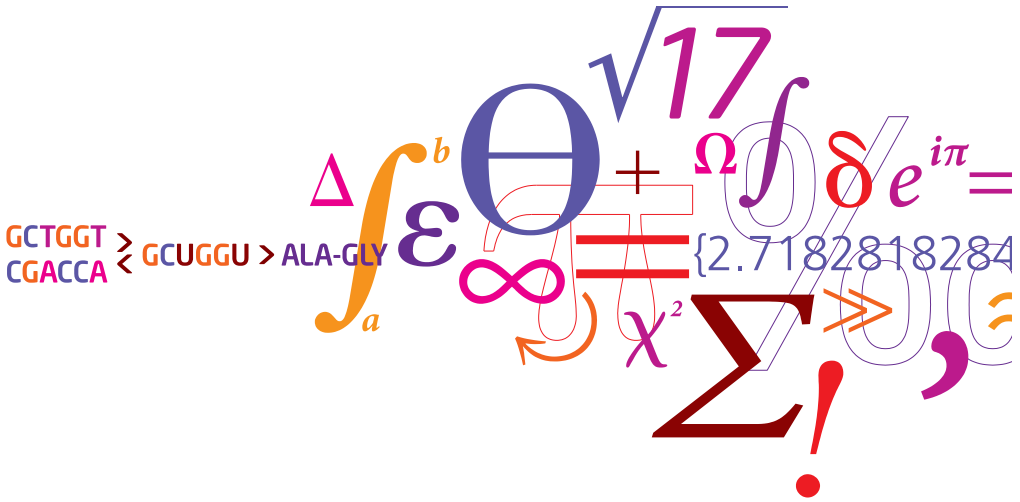
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS
DTU

Basic concepts in Human Variation

Jose MG Izarzugaza (CBS/DTU)



DTU Systems Biology
Department of Systems Biology


Science
THE HUMAN GENOME
nature
the human genome

Variations: Mutations and Polymorphisms

Single Nucleotide Variation (SNV):

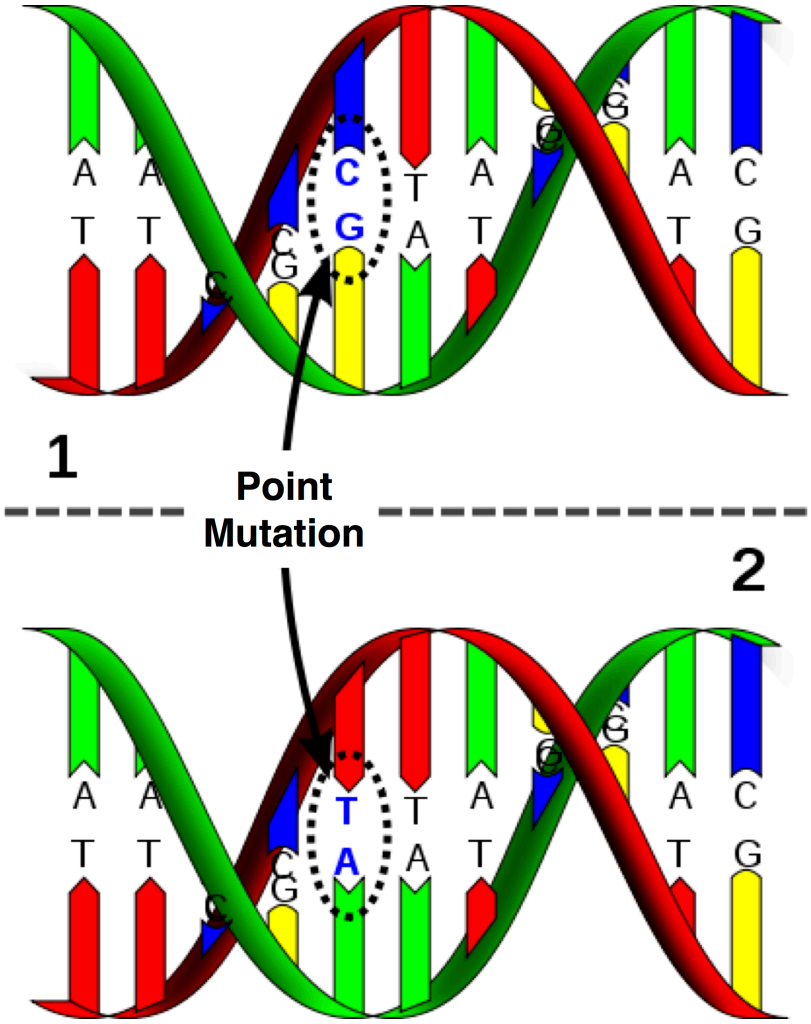
- Permanent change in DNA
- UV, chemicals, tobacco…
- 99.9% identical, $3x10^9$ bp

Polymorphism (SNP):

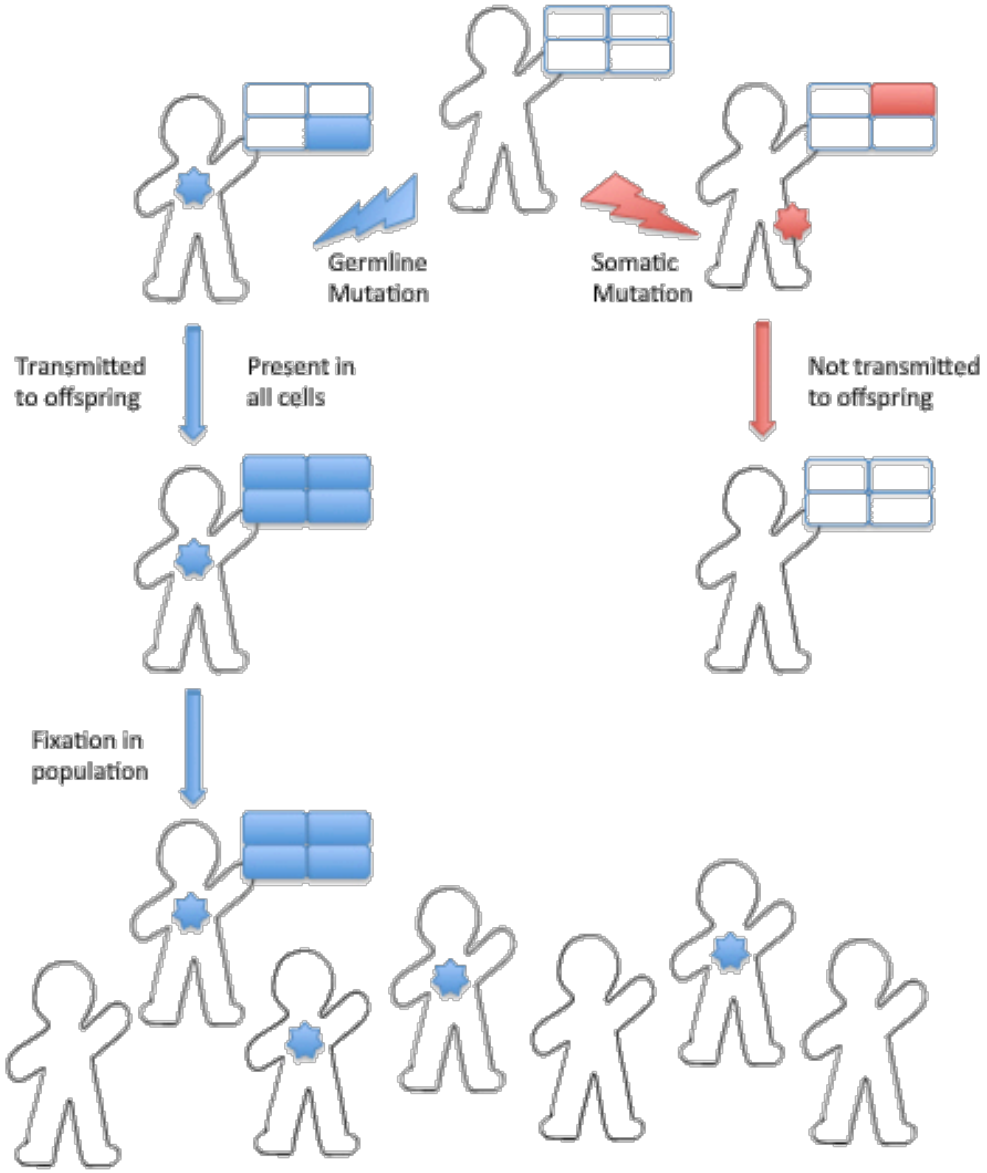
- Implies fixation in population (1%)
- Allele frequencies: Major / Minor

Mutation:

- Usually implies association to disease
- Common allele / Disease allele



Germline *vs* Somatic Mutations



Germline Mutations

- Present in **all** cells
- **Transmitted** to offspring
- **Fixate** in population (SNP)

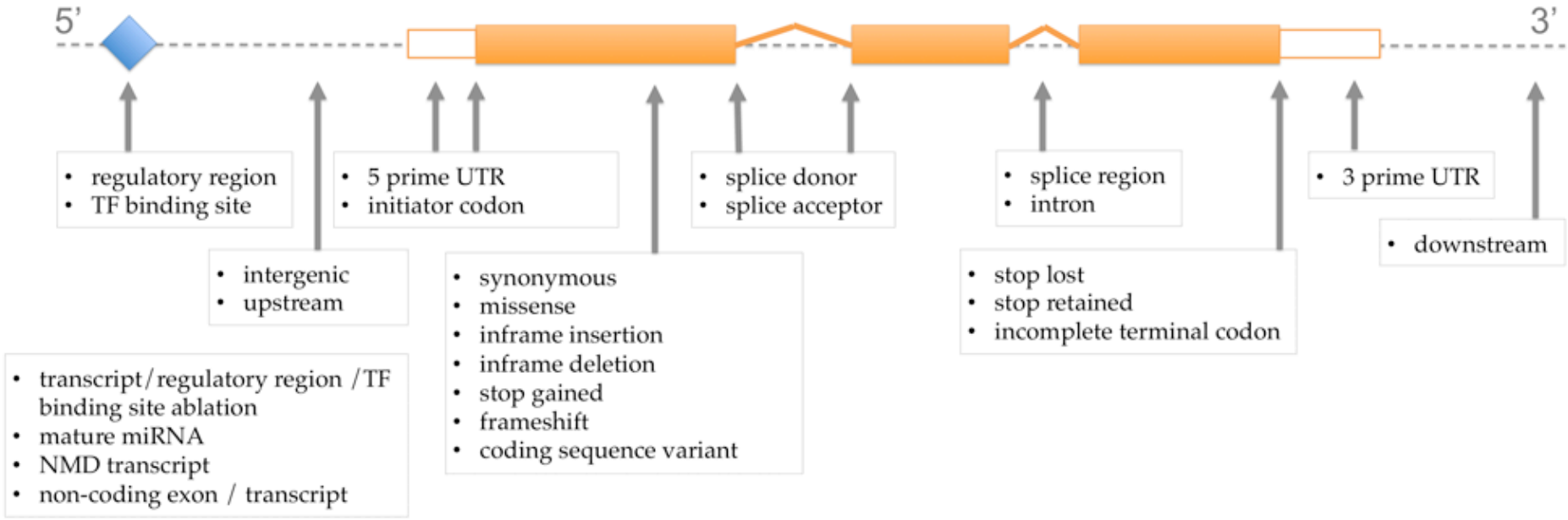
Somatic Mutations

-Present only in **some** cells

-**Not transmitted** to offspring

-Do **not fixate** in population

Coding *vs* Non-Coding variations



Different regions ➔ Different consequences

- Non-Coding mutations: Regulation, transcription, splicing…
- Coding mutations: non-synonymous, synonymous, stop, frameshift…

Synonymous *vs* Non-Synonymous Mutations

Degeneration of the genetic code

1) Synonymous mutation

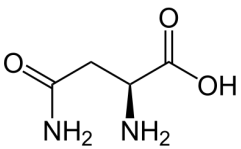
….CCA…. ➔ ….CCC…

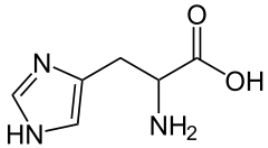
Proline Proline

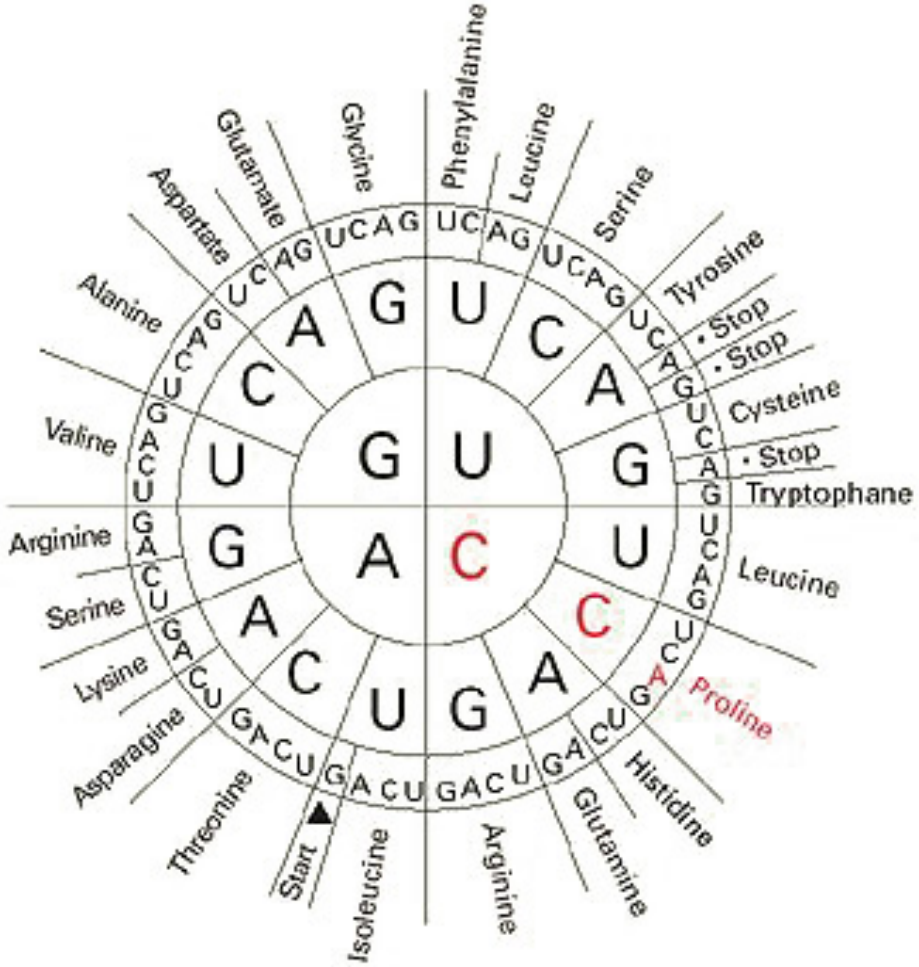
2) Non-Synonymous mutation

….AAC…. ➔ ….CAC…

Asparagine Histidine







Synonymous *vs* Non-Synonymous Mutations

Degeneration of the genetic code

1) Synonymous mutation

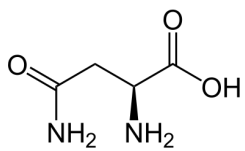
….CCA…. ➔ ….CCC…

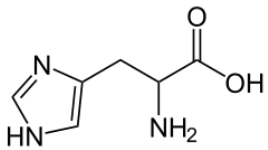
Proline Proline

2) Non-Synonymous mutation

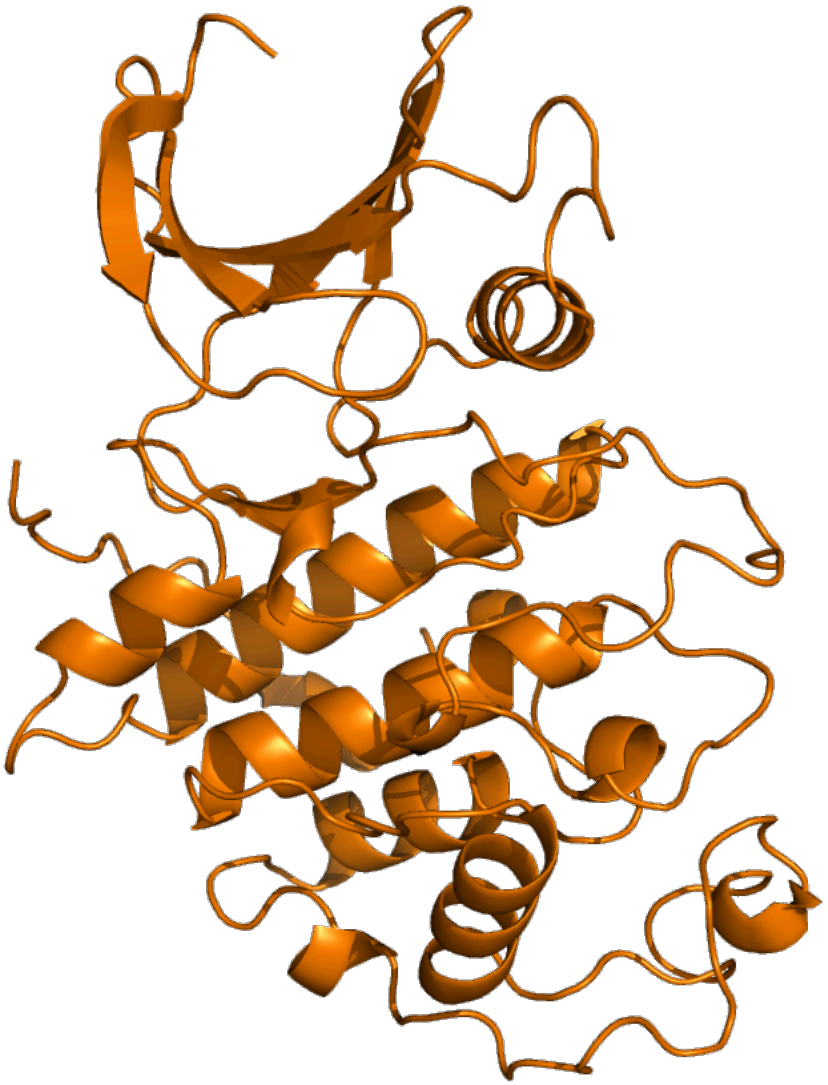
….AAC…. ➔ ….CAC…

Asparagine Histidine

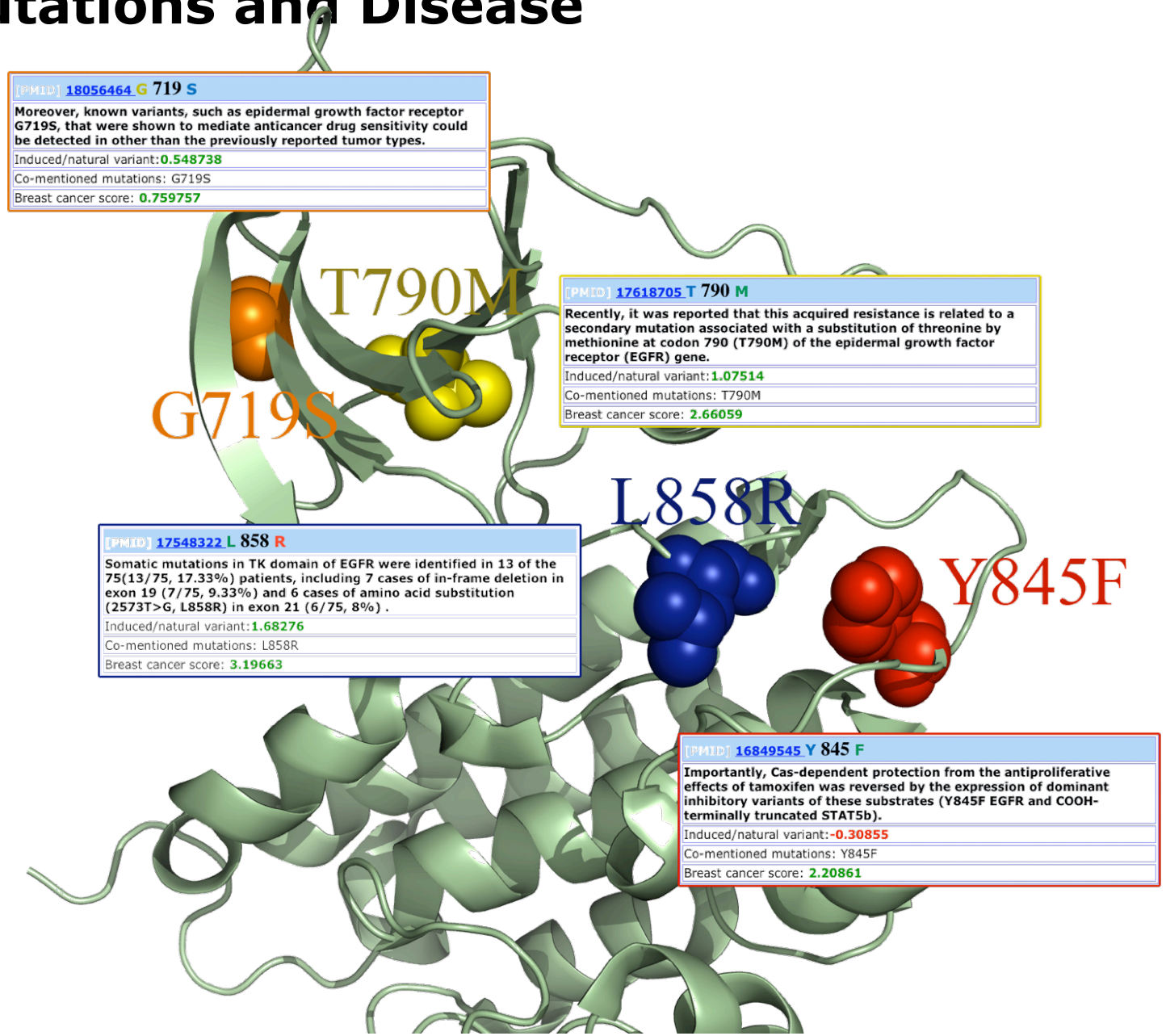




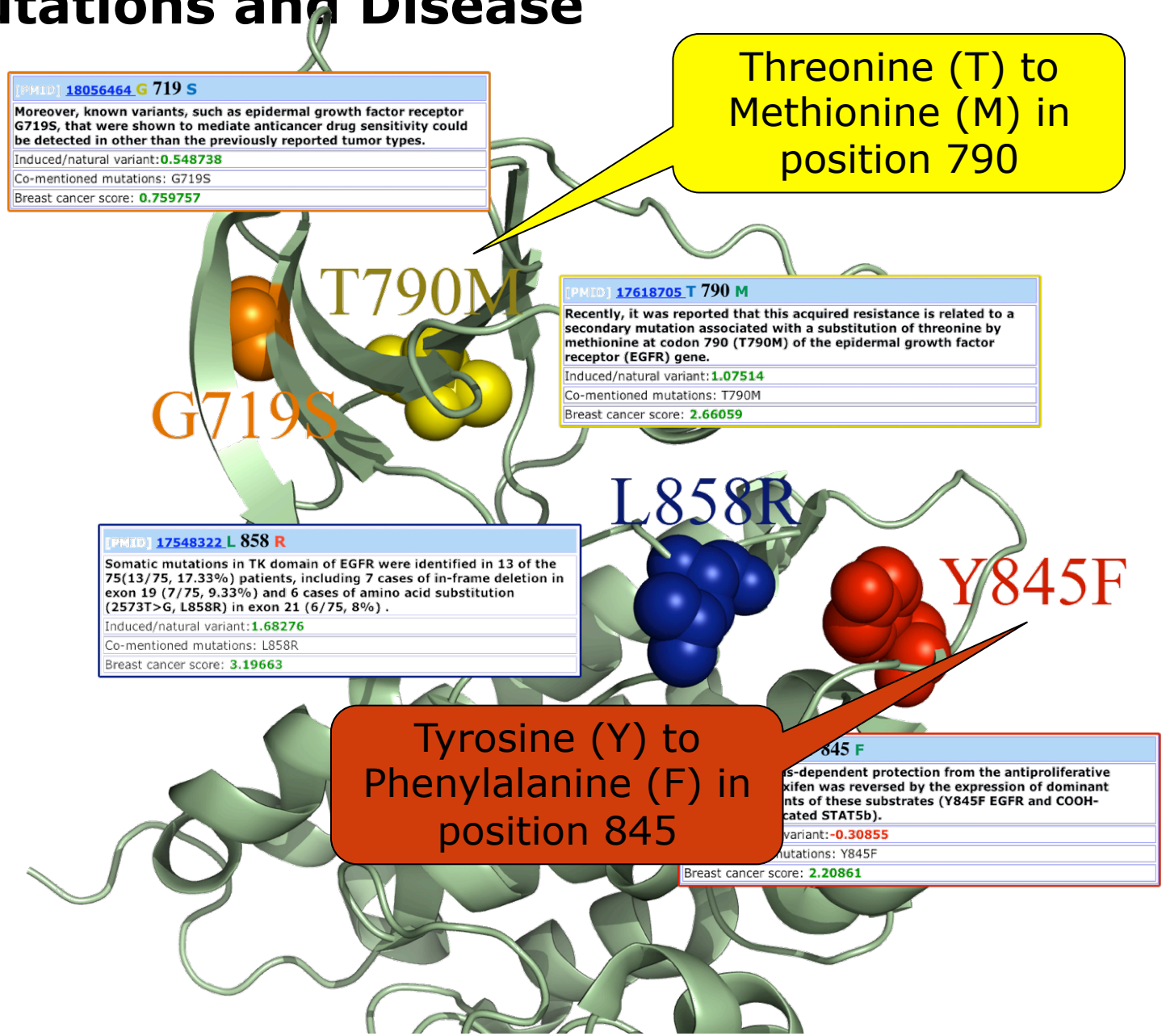
- Different chemical properties
- Altered protein structure



Mutations and Disease

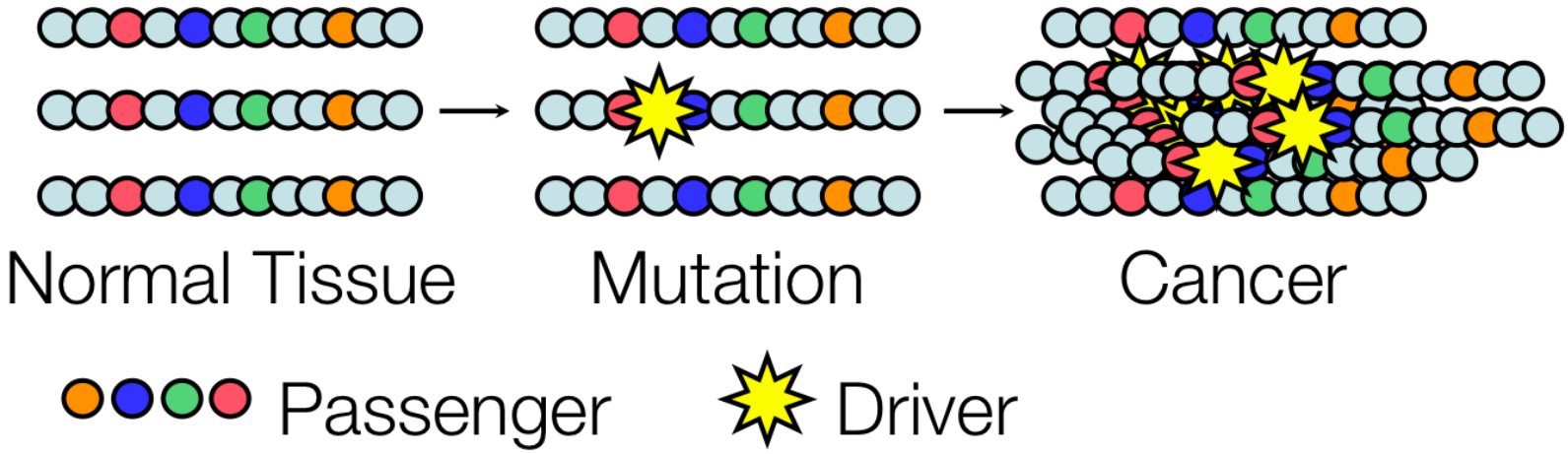


Mutations and Disease



The Drivers and Passengers Paradox

The Drivers and Passengers Paradox



Driver ➔ Confers selective advantage
Disease associated, pathogenic

Passenger ➔ Present in the clonal progenitor
Functionally neutral

Understanding Mutation in -omics times

Traditionally

1 Mutation

=

1 Disease

Phenotype

Function

Mechanism

Lots of hard work

Now (Bioinformatics)

X Mutations

In

Y Patients

And

Z Conditions

Prediction of

Pathogenicity / Unfeasible

Prioritization

SIFT: Pathogenicity of mutations

- SIFT is based on amino acid conservation across species
- Mutation of highly conserved ➔ Pathogenic
- http://blocks.fhcrc.org/sift/SIFT.html
- Pathogenic if **SIFT score < 0.05**

Human	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Chimpanzee	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Orangutan	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Macaque	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Mouse	E	G	R	V	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Rat	E	G	R	V	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Dog	E	S	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Horse	E	G	R	A	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Cow	E	G	R	A	P	D	S	V	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Opossum	E	T	K	T	P	D	S	I	D	Y	R	K	K	G	Y	V	T	P	V	K	N
Zebrafish	V	G	K	L	P	K	S	I	D	Y	R	K	L	G	Y	V	T	S	V	K	N
Medaka	L	I	K	L	P	K	S	V	D	Y	R	K	K	G	M	V	T	S	V	K	N
Tetraodon	V	Q	R	L	P	R	N	L	D	Y	R	K	K	G	A	V	T	A	V	K	D
Tetraodon	G	F	E	T	P	P	S	V	D	W	R	K	A	G	L	V	S	P	V	Q	N
Fugu	G	A	D	L	P	Q	T	V	D	W	R	D	K	G	L	V	T	S	V	K	K

(Ng, 2001)

Methods to predict pathogenicity

2001 – SIFT, SNP3D-stability

2002 – Polyphen

2003 – Panther

2004 – PMUT, PFAM LogRE

2005 – LS-SNP, SNP3D-seq

2007 – SNAP, CanPredict, SAPRED

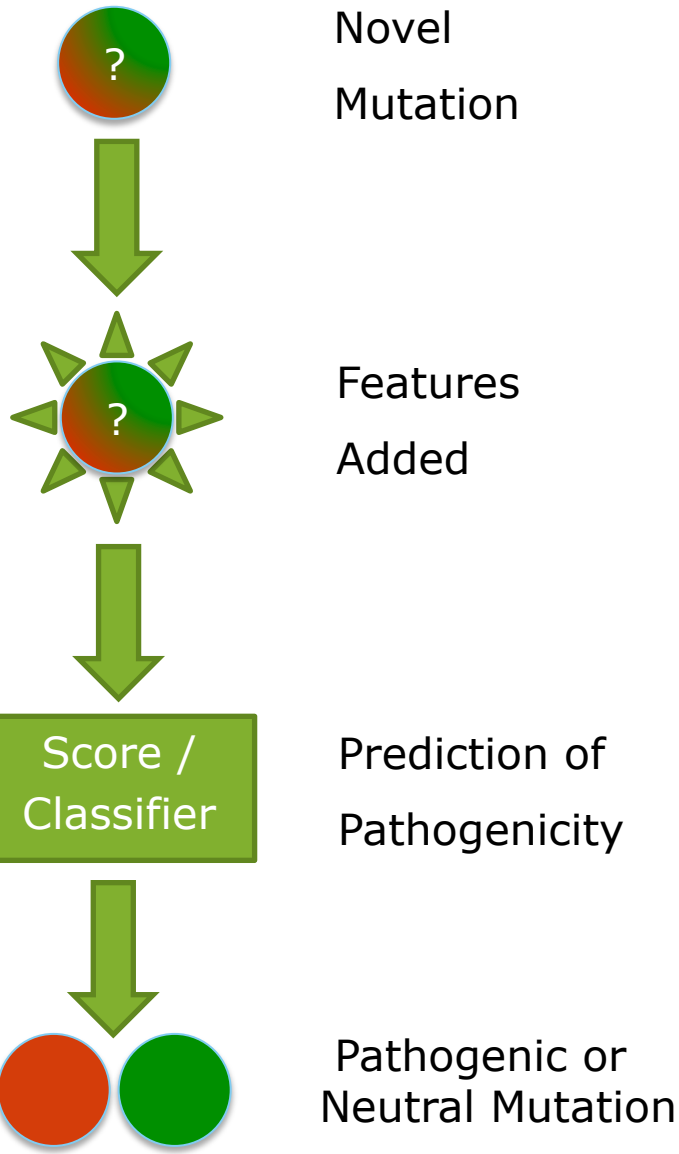
Torkamani (Kinases)

2009 – SNPs&GO

2010 – Polyphen-2, MuD

2012 – KinMut (Kinases)

2013 – NetDiseaseSNP



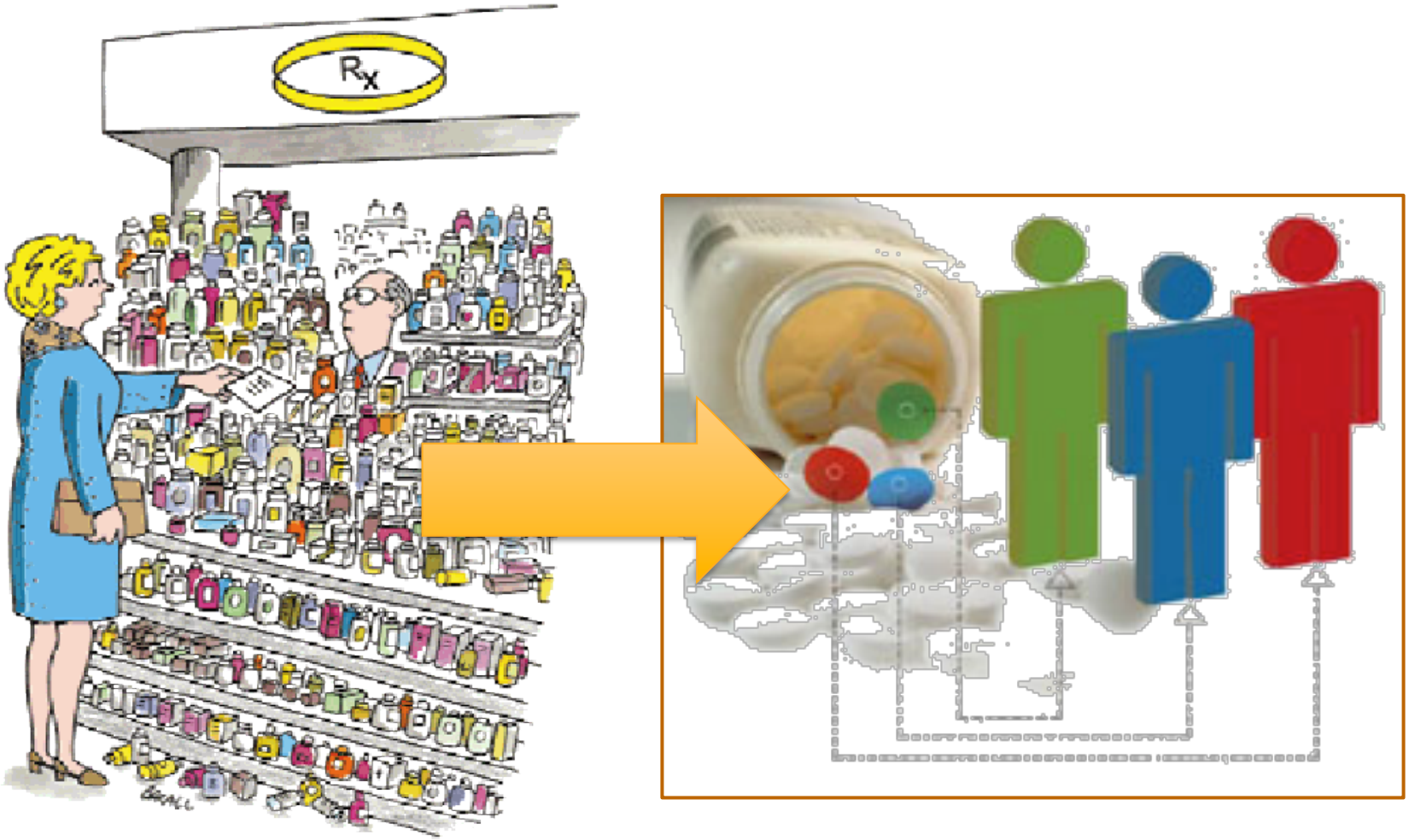
Novel Mutation

Features Added

Prediction of Pathogenicity

Pathogenic or Neutral Mutation

The future: Personalized/Stratified medicine



"Here is my sequence"